Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhanced Micro-Expression Recognition Using MBCC-CNN Architecture

Pratibha Sharma[1], Rajiv Singh[1,2], Swati Nigam*[1,2]
[1]*Department of Computer Science, Banasthali Vidyapith, Rajasthan 304022, India*
[2]*Centre for Artificial Intelligence, Banasthali Vidyapith, Rajasthan 304022, India*
pratibhasjp@gmail.com, jkrajivsingh@gmail.com, swatinigam.au@gmail.com

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

This study presents an innovative approach for micro-expression recognition based on a multi-branch cross-connected convolutional neural network (MBCC-CNN). Unlike conventional single-stream architectures, the proposed framework effectively captures and enhances image features through a multi-branch design integrated with residual connections, Network-in-Network modules, and hierarchical (tree-based) structures. A cross-linked shortcut mechanism is incorporated to merge outputs from different convolutional layers, facilitating smoother inter-branch information flow, expanding the receptive field, and strengthening feature extraction. This design enables efficient fusion of branch-level features, overcoming the limitations of inadequate learning in isolated branches and significantly improving recognition accuracy. Experimental evaluations conducted on benchmark datasets—SAMM, CASME I, CASME II, and CAS(ME)²—achieved recognition rates of 99.89%, 98.73%, 100.00%, and 99.58%, respectively. Comparative analysis with state-of-the-art approaches demonstrates that the proposed MBCC-CNN architecture achieves superior recognition performance and enhanced robustness in micro-expression analysis.

Keywords: MBCC-CNN, N-i-N, FER, SoftMax, ReLU, ResNet

INTRODUCTION

Emotions are an integral component of human existence and are frequently manifested, either consciously or subconsciously, through facial expressions in social encounters. Facial expressions, being a prevalent type of non-verbal communication, are essential for comprehending human emotions [1], which has made them a subject of significant research across multiple domains [2], [3].

Facial expressions are primarily classified into two distinct categories: macro-expressions and micro-expressions (Fig.1). The fundamental difference between them resides in their duration and intensity. Macro-expressions are typically intentional, lasting from 0.5 to 4 seconds, and encompass significant facial muscle movements that affect broader areas of the face. Consequently, they are readily distinguishable from arbitrary facial movements, such as eye blinks. Conversely, micro-expressions are involuntary, nuanced, and ephemeral, often enduring from 0.065 to 0.5 seconds. In contrast to macro-expressions, these cannot be deliberately regulated, therefore frequently revealing concealed emotions that individuals may strive to conceal [5], [6], [8], and [9].

Recognizing micro-expressions is exceedingly difficult due to their ephemeral, involuntary, and low-intensity characteristics. Even highly experienced specialists can reliably identify just approximately 47% of them, and manual analysis is both laborious and susceptible to errors. Micro-expression detection entails identifying the presence of a micro-expression in a video and annotating its onset, apex, and offset. Micro-expression identification involves categorizing identified expressions into emotional classifications, including happiness, sorrow, rage, contempt, or surprise [11].

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Although facial expression recognition (FER) has gained momentum in recent years, it still encounters several hurdles. Factors such as lighting variations, occlusion, and head pose changes, identity differences, and the detection of subtle expressions significantly reduce recognition accuracy. As FER is heavily data-driven, developing robust deep learning models requires large-scale, high-quality datasets, which remain scarce [12]. Traditional handcrafted feature extraction techniques have become inadequate, being prone to noise, incomplete, and inefficient.

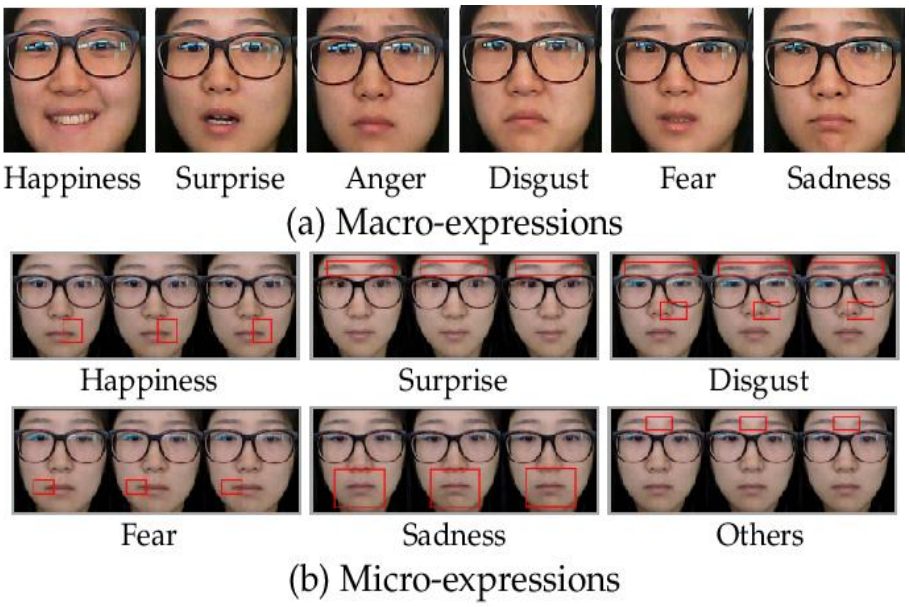

Figure 1: Six macro and micro-expressions from MMEW dataset

Deep neural networks have achieved remarkable outcomes in FER [13]-[18]. Research on facial expressions, originating in the 1970s, was profoundly influenced by Ekman's idea of universal fundamental emotions, which classified seven expressions—happiness, sorrow, anger, fear, surprise, disgust, and contempt—as cross-cultural and universal [22]. Micro-expressions, first defined by Ekman and Friesen [23], are especially relevant in high-stakes situations, where individuals attempt to suppress true emotions due to social or cultural norms. While macro-expressions can be easily recorded with standard cameras, capturing micro-expressions often demands high-speed recording devices, which introduce additional challenges such as noise. Automated recognition of these subtle expressions, initiated only in the late 2000s [24], is therefore a relatively young but promising research area compared to traditional FER.

The contributions of the paper are as follows:

1. Designing an effective CNN-based framework that captures nuanced features of expressions remains an open challenge. To address this, we present a new method termed the Multi-Branch Cross-Connection Convolutional Neural Network (MBCC-CNN).
2. This model enhances feature extraction by incorporating residual connections, cross-connections, Network-in-Network (N-i-N), and a tree-like structure. The cross-connection enables smoother information flow between layers, strengthening feature representation and reducing the risk of insufficient learning.
3. Furthermore, global average pooling reduces parameter size and mitigates overfitting. The obtained features are ultimately classified using a SoftMax classifier to attain precise expression recognition.

This paper is organized as follows: Section 2 categorizes existing facial expression recognition (FER) approaches, Section 3 introduces the MBCC-CNN framework, Section 4 evaluates the model on multiple benchmark micro-expression datasets, and finally, Section 5 concludes the study.

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

LITERATURE REVIEW

Currently, the available approaches for recognizing expressions can generally be grouped into three main categories: techniques that rely on conventional machine learning algorithms, methods built upon convolutional neural networks (CNNs), and hybrid strategies that integrate both traditional approaches and CNN-based frameworks.

2.1. Traditional Machine Learning Methods

In traditional machine learning–based facial expression recognition, manually engineered descriptors such as Gabor wavelet coefficients [26], Local Binary Patterns (LBP) [27], and the widely adopted Histogram of Oriented Gradients (HOG) [28] are commonly employed to characterize expressions.

In [29] Phuc Truong Le Vinh et al. proposed a model on AffectNet and Flickr HQ dataset to evaluates traditional classifiers like Random Forest, SVM and they showed good performance on “happy” as well as “neutral” expression but get in trouble for finding “sad”, “fear”, and “angry” expression.

In [30] Xi Liu came up with overview of traditional ML frameworks in FER that has Decision trees and SVM and contrast them with ANN.it actually gives a structured presentation of traditional algorithm with data pre-processing, preparing the actual model and then training it for FER.

In [31] AMS Gonzalez-Acosta et al. proposed a biometric analysis of emotion recognition using key features of face basically it uses facial landmarks with geometric relationships between them as features and classifiers include MLP, SVM, RDF it signifies that even in the boom of deep learning carefully crafted traditional feature extraction gives nice results.

2.2. Convolutional Neural Network Methods

Traditional approaches for facial expressions often suffer from inefficiency and incomplete representation. As a result, deep learning techniques, particularly those applied to facial expression recognition (FER), have gained increasing attention. In many studies, convolutional neural networks (CNNs) are designed to process facial expression datasets by first performing pre-processing steps and then training/testing the network for emotion classification [35-38].

Kumar et al. [32] brings CNN-based approach for balancing and optimizing data quality, this model shows improved accuracy in finding facial expression by data composition and strategies.

Farooq Akram Alkhaleli et al. [33] proposed a method where CNN models like MobileNet v1 and VGG16 extracts features and then classified by SVM classifier so accuracy get increment by ~3.07% when they merge VGG16 and SVM along with this there is a increment of ~2.74% by combining MobileNet and SVM.

El Boudouri et al. [34] came up with EmoNeXt- tailored ConvNeXt for FER. Which includes channel- wise squeeze and excitation block, self-attention regularizer for feature learning and Spatial Transformer Network to fix the refions for expressions gives better performance on FER2013 dataset.

2.3. Fusion

Despite the considerable success of CNNs in image recognition tasks, issues including low recognition rates, elevated computational cost, and feature degradation continue to afflict FER applications. This research presents a multi-branch cross-connection convolutional neural network (MBCC-CNN) to overcome these constraints. This architecture fundamentally activates residual connections, Network-in-Network modules, and a multi-branch tree configuration. During residual block construction, shortcut cross-connections are added to facilitate smoother information flow across layers. Each branch integrates Network-in-Network modules to strengthen feature extraction across receptive fields, capturing diverse aspects of facial imagery. The extracted features from all branches are then fused, effectively mitigating feature loss. Finally, global average pooling aggregates the last-layer feature maps into a vector, which is fed into a SoftMax layer for classification.

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Rashad et al. [39] proposed an Automated Model (Custom CNN-SVM) for feature extraction with SVM classifier, it has pre-processing steps that has face detection, histogram equalization, and last gamma correction. They took result on multiple datasets (CK+, JAFFE, FER, and KDEF).

In [40], there is a construction of state-of-the-Art for FE classification. It basically follows an approach of evolution of FER from traditional methods (manual feature extraction methods – LBP, HOG, Gabor filters etc.) with deep learning method using CNN (VGG-19, FER-VT and MobileNet-V3), so it bridges between traditional one and new school one, by putting focus on high performance.

Yu et al [41] brings a fusion model that merges face and speech using Global Spatial attention method, it sparkles up feature level fusion with the combination of handcrafted alignment with CNN, basically developed for Affective Behaviour Analysis in Wild (multi-model, multi-scale fusion).

THE PROPOSED METHOD

The method encompasses three primary stages. The face expression dataset undergoes pre-processing to enhance the quality and usability of the input photos. A multi-branch cross-connected convolutional neural network (MBCC-CNN) is subsequently developed to record and extract more nuanced face expression data. The retrieved characteristics are ultimately inputted into a SoftMax classifier, which categorizes the facial expressions into their corresponding classes. Each branch processes features at different receptive fields like Cross-connections share intermediate features across branches - Fusion via concatenation before pooling and classification - Improves robustness to occlusion and captures multi-scale cues (Fig. 2).

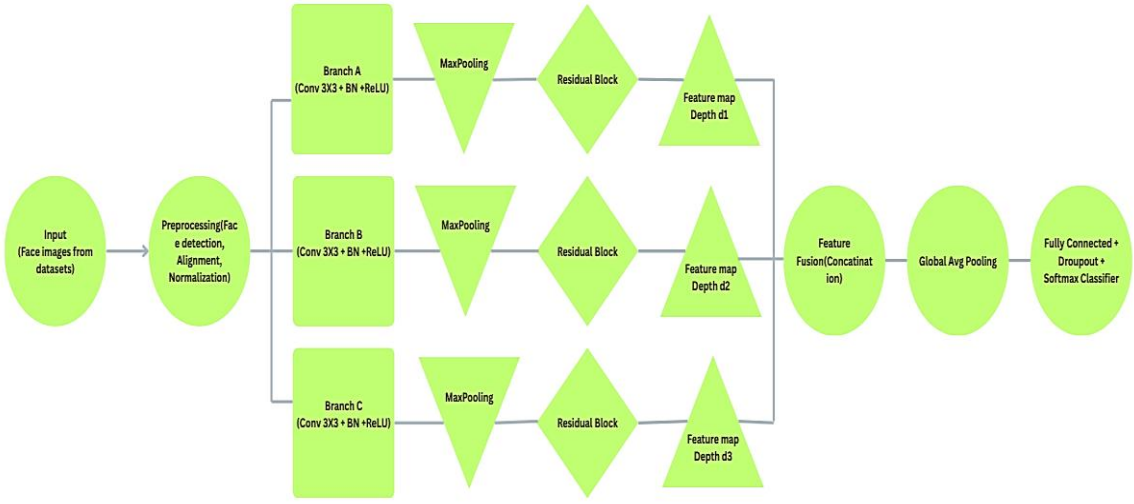

Figure 2: Pipeline flow for MBCC-CNN model

3.1. Input images

Micro-expression datasets are short clips with subtle facial motion, this step supply the model with a normalized, fixed-size representation, Channels either grayscale (texture matters) or RGB if colour helps; for motion branches you will feed optical flow or frame-difference maps. These input images serve as raw data for the network and typically include a variety of facial expressions, poses, lighting conditions, and backgrounds. Grayscale image of size 64×64, represented by the shape (64, 64, 1).

3.2. Preprocessing

The process begins by identifying and locating the face within the input image with the Haar Cascade technique, which employs rectangular Haar features to discern differences between the eyes and cheeks, subsequently calculating the sum of pixels to rapidly compute features. The MTCNN algorithm employs a staged approach: the Proposal Network scans the image at various scales to suggest candidate face regions, the Refine Network enhances

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

these proposals by eliminating false positives, and the Output Network identifies facial landmarks. Conversely, the Dlib algorithm primarily integrates the methodologies of SVM and HOG.

Once the face is detected, the facial landmarks are used to align the face into a standard pose. This normalization reduces the mutability caused by head tilts or rotations then Pixel values of the image are normalized (e.g., scaling to [0, 1] or [-1, 1]) to reduce the effects of lighting variations and to ensure consistent input for the convolutional layers. The output of this pre-processing block is a set of uniform, aligned, and normalized face images that are directly moved into the multi-branch convolutional for feature extraction.

3.3. Network Structure of MBCC-CNN model

It is designed for image-based classification, possibly for tasks such as FER, emotion detection, or fine-grained image analysis. The model integrates several modules, each focusing on learning different levels and types of features, and it applies feature fusion and attention-based classification for the final output (Fig. 3).

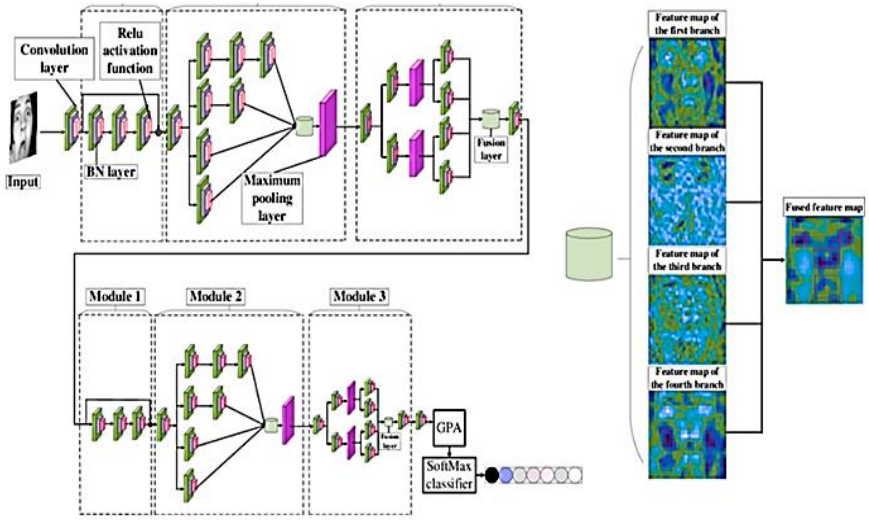

Figure 3: Multi-branch deep CNN architecture

3.3.1. Structure of Module 1

This module adopts the residual connection strategy [42], wherein the input is directly bypassed to the output, allowing the network to learn solely the difference between the input and output, hence simplifying feature learning. ResNet method and feature fusion method plays vital role in it, the main motto is to improve gradient flow, ease the training of model and to expend the learned features deeply and diversely (Fig. 4).

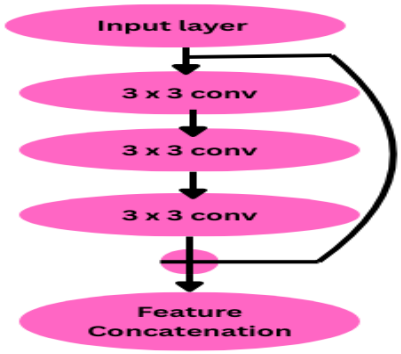

Figure 4: Residual convolutional block with feature concatenation

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.3.2. Structure of Module 2

This module opts the Network-in-Network aspect for feature extraction and enhances the ability of discrimination. When it gets compared with single network then it finds that multiple branch network can extract abstract features of different channels efficiently then it coupled them to improve extraction ability. This module is specially designed to use size of convolutional filters and pooling operations parallel to extract multi-scale features. After this these features get fused by concatenating them (Fig. 5).

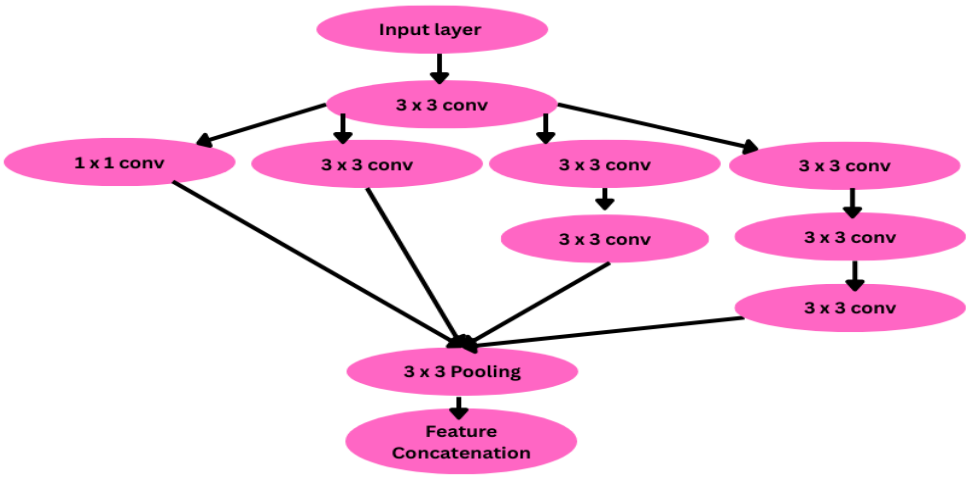

Figure 5: multi-branch convolutional module

3.3.3. Structure of Module 3

This module seems tree like structure with multi branch structure and these branches needs small amount of data and at the backend convolutional layer and maximum pooling layer is used to form hierarchical structure on the other hand it has N-i-N which helps in extracting features of image efficiently.So, branch network structure can fetch features of different images and then merge them, by this feature extraction efficiency get improved of network, the kernel size is 3 × 3 for the layers. After a layers of convolutional the input image gets divided into different branches so that features can extracted deeply and nicely, then they coupled so that loss of useful features get decreased and this process improves the recognition performance of network (Fig 6).

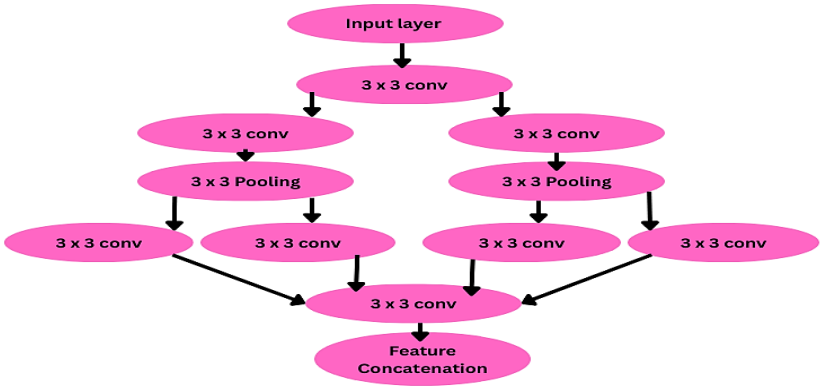

Figure 6: MB-CNN module incorporates parallel convolutional and pooling operations, followed by feature fusion via concatenation

3.4. Fusion of Branch features

For module 2 and 3, feature fusion important for addition, without changing number of channels. Assume, that the image size of the input module is W × W × D, then image size becomes $W_1 \times W_1 \times D_1$ after the application of

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

convolution layer. Then, the size of the image is $W_1 \times W_1 \times D_1$ after the pooling layer with $F_1 \times F_1$ as the pooling core and S_1 as the step size. In this case, it is assumed that the four input channels X_i, Y_i, Z_i and M_i where i = 1, 2, 3, . . . n. For the addition process, the fused feature is recorded as N_{add}

$$W1 = \frac{W+(2*P)-F}{S} + 1 \quad (1)$$

$$D1 = K \quad (2)$$

$$W2 = \frac{W1-F1}{S1} + 1 \quad (3)$$

$$D2 = D1 \quad (4)$$

$$Nadd = \sum_{i=1}^{n} (Xi + Yi + Zi + Mi) * Ki = \sum_{i=1}^{n} Xi * Ki + \sum_{i=1}^{n} Yi * Ki + \sum_{i=1}^{n} Zi * Ki + \sum_{i=1}^{n} Mi * Ki \quad (5)$$

Where,

K = number of convolution kernels

F = size of convolution kernels

S= step size

P = zero padding

Ki = number of convolution kernels corresponding to i

'∗' = convolution symbol.

Figure 7 represents the architecture of a Multi-Branch Concatenated Convolutional Neural Network (MBCC-CNN) majorly designed to process two separate inputs from different sources and after that they got merged for further classification, where each branch is designed to independently process separate input data. Each input is a grayscale image of size 64×64, represented by the shape (64, 64, 1).

In the Ist branch, the image gets passed through a two-layer convolutional block in that first convolutional layer applies 32 filters, conserving spatial dimensions after that it has MaxPooling2D which reduces the size of feature map in half. The output passed through a second convolutional layer consist of 64 filters, again followed by a MaxPooling2D layer, which further reduces in spatial dimensions. The final feature map is flattened into a one-dimensional vector of size 16,384.

The 2nd branch twins exactly with first branch. It processes a second input image of the same shape and applies the same sequence of convolutional and pooling operations. Like the first branch, the output is also flattened into a vector of size 16,384.

So, outputs of both branches concatenated to form combined feature vector of size 32,768. This merged representation passed to fully connected dense layer with 128 neurons, helps in non-linearity and enabling the model to learn abstract feature representations. A Dropout layer is employed next to reduce overfitting by randomly disabling a fraction of neurons during training.

Finally, the model includes a second dense layer with 20 output neurons, which act as the output layer. This layer does the final classification, using a softmax activation function, assuming a multi-class classification problem involving 20 distinct classes.

The architecture is designed to extract hierarchical features from two separate inputs and merge them effectively for sturdy classification. The dual-branch design, combined with concatenation and fully connected layers, allows the network to capture complementary information and improves overall performance.

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

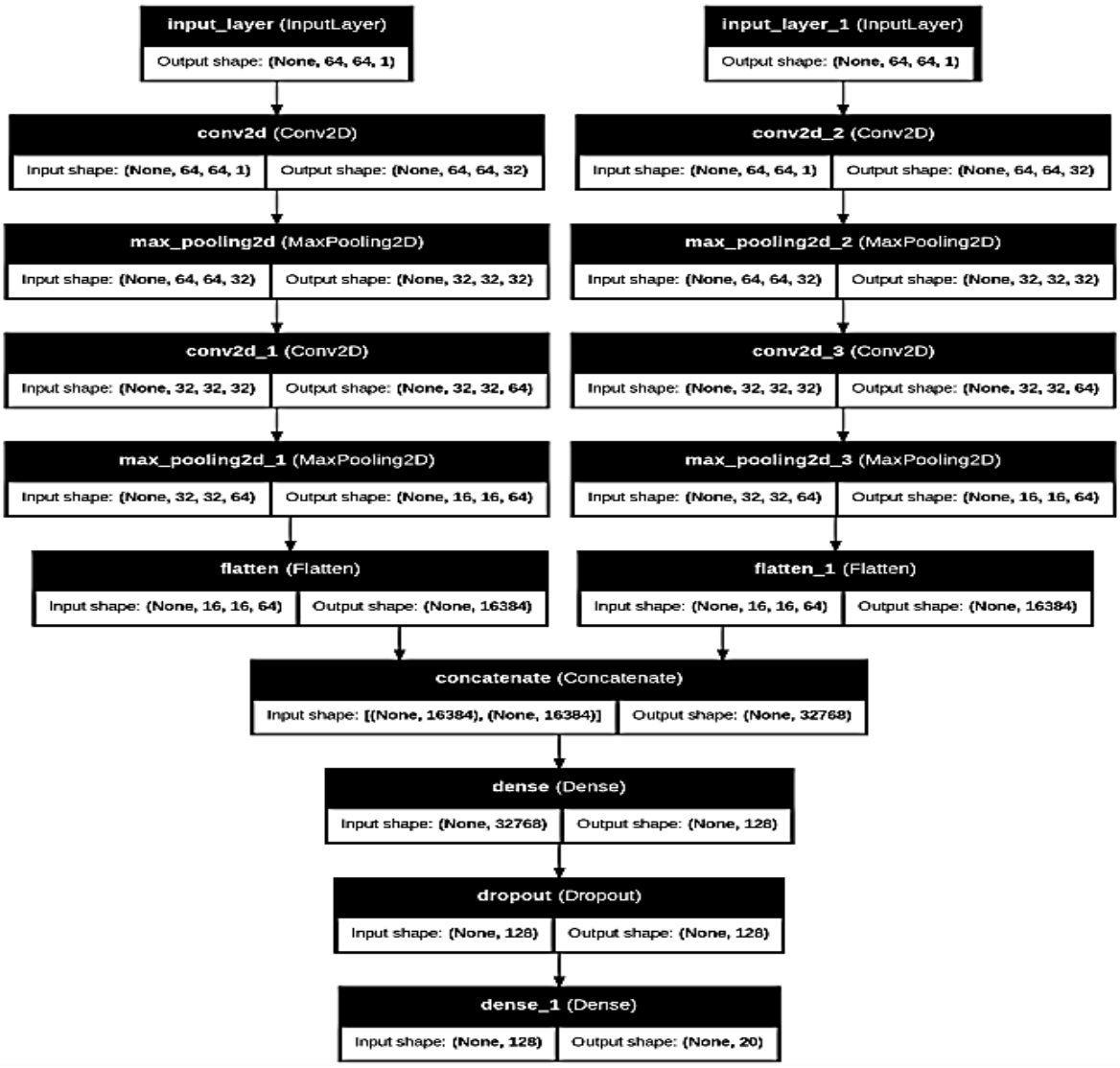

Figure 7: Dual-branch CNN architecture

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The proposed Multi-Branch Cross-Connection Convolutional Neural Network (MBCC-CNN) is designed to simultaneously capture both local and global discriminative features for sturdy facial expression recognition. The architecture employs a dual-branch strategy, where two independent convolutional pathways process the same input image in parallel. The first branch applies convolutional layers with 3×3 kernels, which are efficient in extracting fine-grained texture information and subtle local variations present in facial regions such as the eyes, mouth, and eyebrows. In contrast, the second branch utilizes larger 5×5 kernels, enabling the network to learn broader spatial dependencies and global structures that contribute to the holistic representation of facial expressions. Each branch is composed of two convolutional layers with 32 and 64 filters, respectively, followed by max-pooling operations with a 2×2 window, which downsample feature maps to reduce computational complexity while preserving salient features. After convolution and pooling, the feature maps in each branch are flattened into one-dimensional vectors, ensuring compatibility for subsequent fusion.

The outputs of both branches are concatenated to form a unified feature representation, effectively integrating localized and global-level information within a single embedding space. This fusion mechanism enhances the expressive power of the network by capturing complementary features from different receptive fields. The merged representation is passed through a fully connected dense layer with 128 neurons activated by the ReLU function, providing a high-capacity feature transformation layer for nonlinear abstraction. To mitigate overfitting and improve generalization, a dropout layer with a probability of 0.5 is applied. Finally, the network concludes with a softmax layer, which outputs a probability distribution across the predefined emotion categories.

For training, the MBCC-CNN model is compiled using the Adam optimizer with an initial learning rate of 0.001, ensuring efficient gradient updates with adaptive moment estimation. The categorical cross-entropy loss function is employed, as it is well-suited for multi-class classification tasks, and model performance is assessed in terms of classification accuracy. The multi-branch design, with kernels of varying sizes, ensures that the network effectively balances local detail preservation with global context extraction, thereby improving the robustness of facial expression classification across diverse datasets.

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Spontaneous Actions and Micro-Movements (SAMM) Dataset Results

The Spontaneous Actions and Micro-Movements (SAMM) dataset [65] represents the first high-resolution collection focused on micro-expressions, consisting of 159 spontaneously triggered facial micro-movements and encompassing a wide range of demographic variations. The elicitation process relied on the 7 fundamental emotions [66] and was captured at a frame rate of 200 fps.

Unlike traditional approaches where self-reports are gathered after exposure, the SAMM dataset has specific images for each class. In total, seven types of classes were employed to evoke emotional responses, with participants instructed to suppress their outward expressions—thereby encouraging the appearance of micro-movements.

Figure 8: Sample of SAMM dataset with anger expression

To intensify the likelihood of suppressed expressions, a monetary incentive of 50 was awarded to the participant judged most successful in concealing their feelings, creating a high-pressure scenario [66], [67]. Prior to participation, every subject filled out a pre-experiment questionnaire, enabling researchers to adjust stimuli based on individual differences and enhance emotional engagement. Altogether, the dataset reports 159 FACS-annotated micro-movements (Fig. 8).

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Detailed analysis of the comparison table for the SAMM dataset that includes performance (accuracy) of various state-of-the-art methods used for micro-expression recognition. The table shows significant differences in performance (see table 1). The training and validation accuracy curves of a model achieving a final accuracy of ~99.89%. X-axis has Epochs (0 to 20) and Y-axis has Accuracy (range: 0.6 to 1.0) Blue line represents Training accuracy (train acc) and orange line shows Validation accuracy (val acc). Training accuracy starts at around 61% at epoch 0. Validation accuracy starts around 79%. Both training and validation accuracies increase sharply in the first 5 epochs. Validation accuracy reaches around 98% by epoch 5. Training accuracy reaches around 95% by epoch 5. After epoch 5, both curves show steady improvement. Validation accuracy peaks near 100% around epoch 15. Training accuracy stabilizes around 98-99% near epoch 15 and remains steady afterward (See fig. 9) and the loss curves for the MBCC-CNN model trained on the SAMM, showing how both training and validation loss evolves across epochs. The final loss reaches ~0.01, indicating very high model confidence and minimal error. X-axis has Epochs (0 to 20) and Y-axis has Loss (range ~0 to 1.1) so, blue line shows Training loss (train loss) along with orange line represent Validation loss (val loss). Training loss starts around 1.1 at epoch 0. Validation loss starts lower, near 0.5 at epoch 0. Both training and validation losses drop sharply within the first 5 epochs. Validation loss drops faster initially; falling below 0.1 by around epoch 7. Both losses stabilize after approximately epoch 10. Training loss settles around 0.03. Validation loss converges slightly lower, around 0.015. (See fig. 10), the confusion matrix for a classification model tested on the SAMM.

The confusion matrix summarizes the model's performance across 8 different classes. X-axis shows Predicted class labels (0 to 7) and Y-axis show True class labels (0 to 7). From light blue (low values) to dark blue (high values), indicating frequency. For class 0 correct predictions (Diagonal) is 411 and 1 is Misclassifications, class 1 correct predictions (Diagonal) is 180 and no Misclassifications, class 2 correct predictions (Diagonal) is 47 and no Misclassifications, class 3 correct predictions (Diagonal) is 84 and no Misclassifications, class 4 correct predictions (Diagonal) is 387 and 1 is Misclassifications, class 5 correct predictions (Diagonal) is 85 and no Misclassifications, class 6 correct predictions (Diagonal) is 212 and no Misclassifications and last for class 7 correct predictions (Diagonal) is 403 and 0 is Misclassifications (Fig. 11).

The Receiver Operating Characteristic (ROC) curves for multiple classes in the SAMM dataset. X-axis (False Positive Rate) from 0 to 1, Y-axis (True Positive Rate) from 0 to 1. The diagonal dashed black line represents a random classifier (AUC = 0.5), used as a baseline. Anger, Contempt, Disgust, Fear, Happiness, Sadness, Surprise, Other All these have AUC = 1.00, which means 100% true positive rate at all false positive rates. The model perfectly distinguishes each class from the others (Fig. 12).

Table 1: Comparisons for the SAMM dataset

Methods	**Accuracy (%)**
FaceSleuth (SOA + CVA) [44]	87.1
FACS-Based Graph Features [46]	87.33
AUMEs (AU Detection-Based Dual-Stream Multi-task 3DCNN) [47]	79.85
OC-Net with GAN Data Augmentation [48]	71.72
OFF-ApexNet [49]	74.60
STSTNet (Shallow Triple Stream 3D CNN) [50]	76.05
MERSiamC3D [51]	64.03
Vision-Transformer variant (convolutional patch embeddings + dual/attention branches) [61]	100.00
Proposed	**99.89**

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

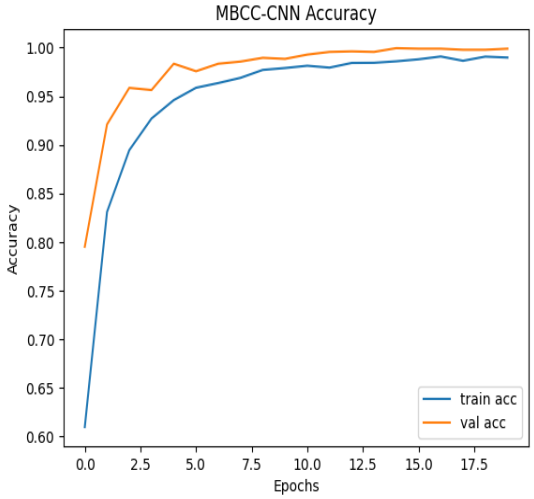

Figure 9: Training and validation accuracy curves for SAMM dataset

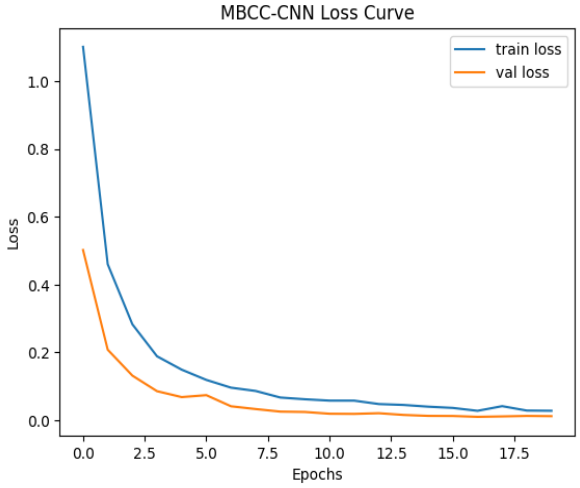

Figure 10: Loss curves for the MBCC-CNN model trained on the SAMM dataset

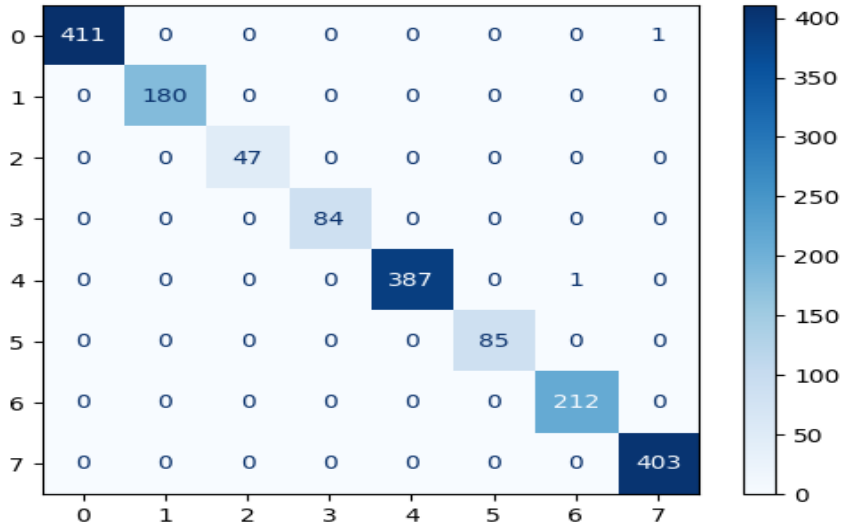

Figure 11: Confusion matrix for a classification model tested on the SAMM dataset

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

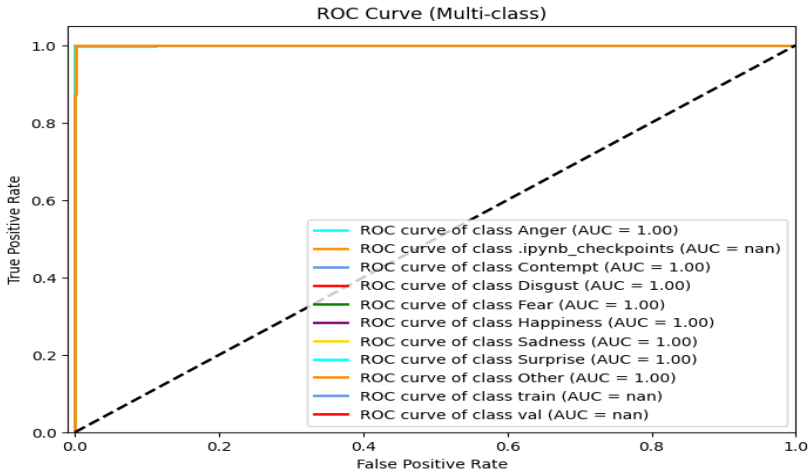

Figure 12: ROC curve for multiple classes in SAMM dataset

4.2 Chinese Academy of Sciences Micro-Expressions (CASME I)

Yan et al. [69] developed a spontaneous micro-expression dataset named CASME. This dataset includes 195 micro-expression video samples captured at a frame rate of 60 frames per second (fps). These 195 clips were carefully chosen from over 1,500 facial movement recordings, involving 35 subjects (13 women and 22 men). Detailed analysis of the comparison table for the CASME I dataset that includes performance (accuracy) of various state-of-the-art methods used for micro-expression recognition. The table shows significant differences in performance (see table 2). The training and validation accuracy curves of a model named MBCC-CNN trained on the CASME I dataset, achieving a final accuracy of ~98.73%. X-axis (Epochs) from 0 to 20, Y-axis (Accuracy) from 0.70 to 1.00. Blue line resembles Training Accuracy (train acc) and orange line shows Validation Accuracy (val acc). Starting Point of training accuracy starts Around 0.73 at epoch 0 but had Sharp rise between epoch 0 and epoch 2, reaching ~0.95 and stabilize From epoch 3 onward, training accuracy fluctuates slightly around 0.965–0.975, indicating convergence. So, Final Value is 0.975. Validation accuracy starts around 0.975, higher than training accuracy even in early epochs. Stable stays between 0.975–0.99 throughout training. Final Value is Closes near 0.99, indicating excellent generalization. (see fig. 14) and the loss curves for the MBCC-CNN model trained on the CASME I, showing how both training and validation loss evolves across epochs. The final loss reaches ~0.02, indicating very high model confidence and minimal error. X-axis we applied epochs Ranges from 0 to 20, Y-axis (Loss) has Ranges from 0 to approximately 0.95. Blue line indicates Training Loss and Orange line indicates Validation Loss. Initial Value is Very high (~0.93), indicating large error at model initialization but Sharp Decrease significantly in the first few epochs (1–3), suggesting rapid initial learning. Stabilize From epoch 4 onwards, the loss stabilizes around ~0.06 to ~0.08 so, Final Value ~0.05 by epoch 20, indicating low training error. Initial Value: Starts around ~0.08. Stability: Validation loss quickly drops to ~0.03 by epoch 2 and remains relatively stable throughout. Final Value: Ends around ~0.025, consistently lower than training loss. (See fig. 15), the confusion matrix for a classification model tested on the CASME I. The confusion matrix summarizes the model's performance across different classes. 19 × 19 (indicating classification across 19 classes) Colour Bar: Right-hand side shows intensity (dark blue = high count), All non-diagonal elements are 0 this implies 100% accuracy, precision, recall, and F1-score across all classes.Class 6 has 102 samples (most frequent) and Class 4 has 9 samples (least frequent), Despite class imbalance, the model shows perfect classification across all classes (Fig. 16).

The Receiver Operating Characteristic (ROC) curvesfor multiple classes in the CASME I dataset. ROC curves are commonly used to evaluate the performance of classification models by illustrating the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across different threshold settings. X-axis shows False Positive Rate (FPR), Y-axis shows True Positive Rate (TPR). Each curve represents a separate subject class (e.g., sub01, sub02... sub19), 17 out of 19 subjects: AUC = 1.00 but 2 subjects (sub10, sub11) have AUC = 0.99, Most AUC values are 1.00, which means the classifier perfectly distinguishes those subjects' micro-expressions. For sub10 and sub11, AUC = 0.99, still indicating very high separability. (Fig. 17).

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The dataset is categorized into two groups based on the recording environment and the type of camera employed.

- Category A: Recordings in this group were obtained using a BenQ M31 camera operating at 60 fps, with a resolution of 1280 × 720 pixels. The sessions were recorded under natural lighting conditions.
- Category B: Samples in this group were captured with a GRAS-03K2C camera, also at 60 fps, but with a resolution of 640 × 480 pixels. For this setup, two LED light sources were used during recording.

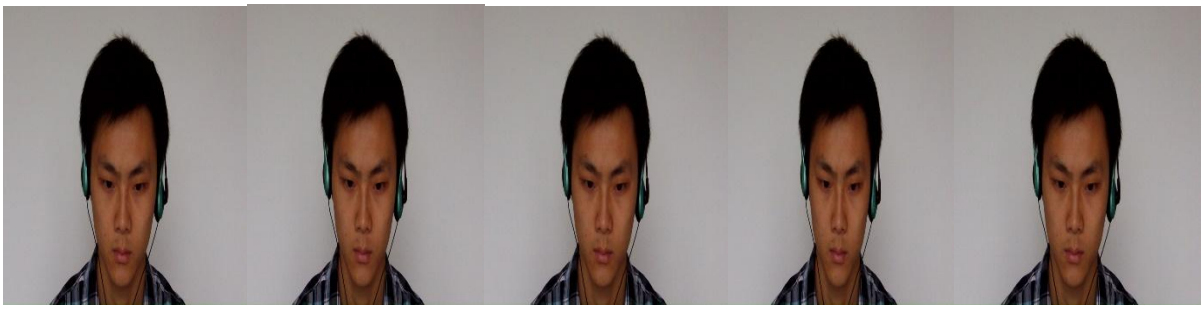

Figure 13: Sample images of CASME I dataset

Table 2: Comparisons for the CASME I dataset

Methods	Accuracy (%)
LEARNet (Lateral Accretive Hybrid Network) [56]	80.62
DSTAN (Dual-Stream Spatiotemporal Attention Network) [57]	77.91
Automatic Micro-Expression Recognition Using LBP-SIPl and FR-CNN [58]	81.56
Enhanced Local Cube Binary Pattern (handcrafted spatio-temporal descriptor) [59]	88.70
3D residual attention network (spatio-temporal + attention) [60]	90.5
Vision-Transformer variant (convolutional patch embeddings + dual/attention branches) [61]	96.00
Proposed	**98.73**

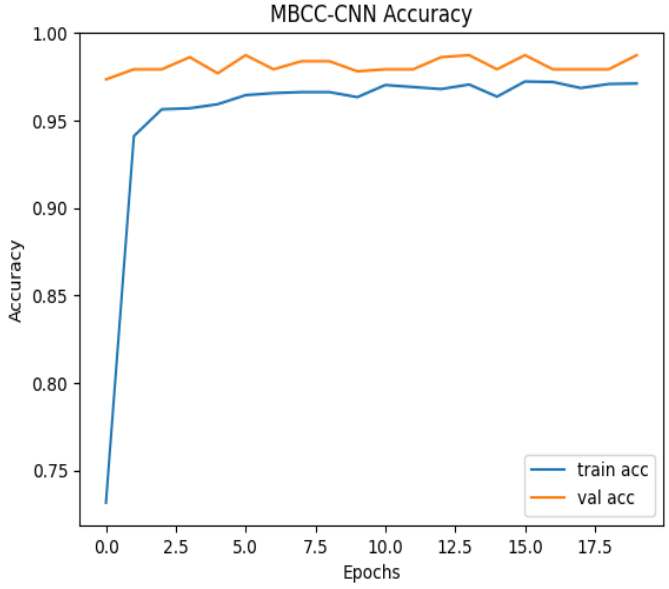

Figure 14: Training and validation accuracy curves for CASME I dataset

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

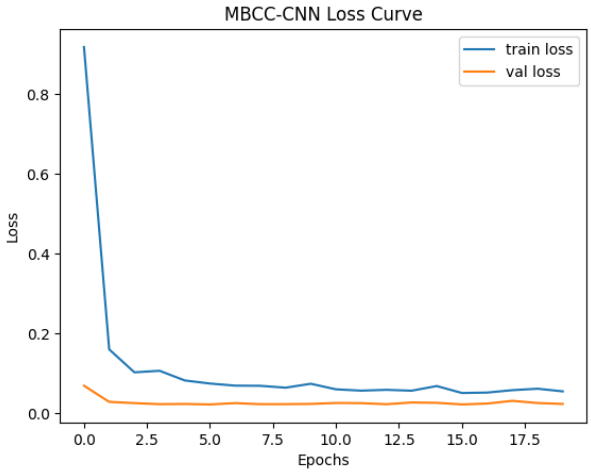

Figure 15: Loss curves for the MBCC-CNN model trained on the CASME I dataset

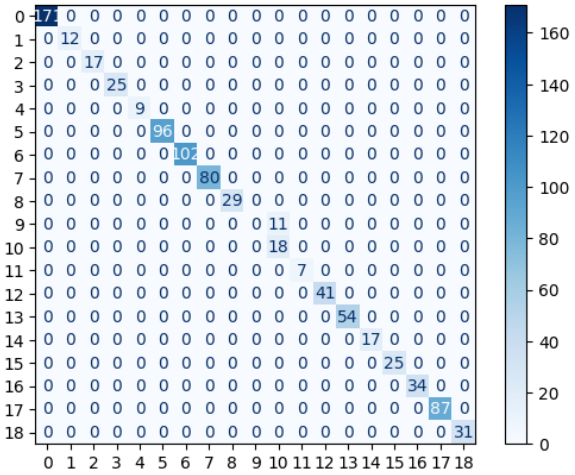

Figure 16: Confusion matrix for a classification model tested on the CASME I dataset

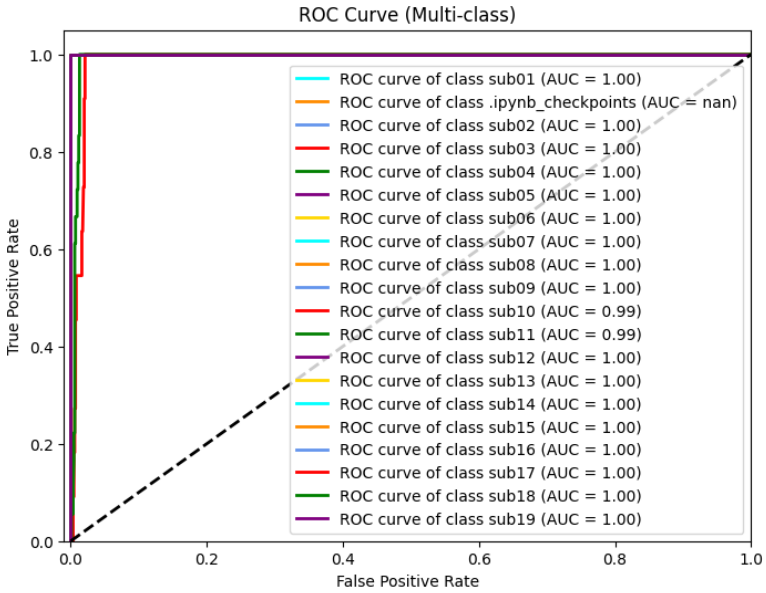

Figure 17: ROC curve for multiple classes in CASME I dataset

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.3 Chinese Academy of Sciences Micro-Expression II (CASME II) Dataset Results

Yan et al. [70] introduced the CASME II dataset as an enhanced successor to CASME [69], bringing notable advancements. All recordings in CASME II consist of natural, dynamic micro-expressions captured at a high temporal resolution of 200 frames per second. To support detection studies, each sequence includes a few frames both before and after the micro-expression, though the number of these surrounding frames differs across samples. The raw video resolution is 640 × 480 pixels, later compressed into MJPEG format, with cropped facial regions sized around 280 × 340 pixels. The micro-expressions were obtained under carefully controlled laboratory settings. In total, the dataset comprises 247 micro-expressions collected from 35 individuals, selected from nearly 3000 facial movements, and annotated with Action Units (AUs) in accordance with the Facial Action Coding System (FACS) [71]. Care was taken during recordings to eliminate lighting flickers and to minimize reflections on facial areas [68] Detailed analysis of the comparison table for the CASME II dataset that includes performance (accuracy) of various state-of-the-art methods used for micro-expression recognition. The table shows significant differences in performance (see table 3). The training and validation accuracy curves of a model named MBCC-CNN trained on the CASME II dataset, achieving a final accuracy of 100.00%.X-axishasNumber of epochs (0 to 19), Y-axis has Accuracy (0.60 to 1.00), Blue line indicates Training accuracy over epochs, Orange line indicates Validation accuracy over epochs, The training accuracy starts at ~60% in the first epoch and jumps above 95% by the second epoch.Indicates the model learns quickly and effectively extracts discriminative features early. Training accuracy fluctuates slightly but remains in the range of 96–99% throughout and Validation accuracy is nearly perfect (≈1.0) across all 20 epochs (Fig. 19) and the loss curves for the MBCC-CNN model trained on the CASME II, showing how both training and validation loss evolves across epochs, indicating very high model confidence and minimal error. X-axis has Epochs (0–19), Y-axis has Loss (presumably cross-entropy or similar)Blue line shows Training loss, Orange line indicates Validation loss, at epoch 0, training loss starts high (~1.45), indicating model begins with random weights.Sharp drop in the first few epochs, suggesting rapid learningand effective initial training.Validation loss starts low (~0.03) and remains nearly flat andminimal(~0.01 to 0.02).After ~5 epochs, both training and validation losses plateau.Final training loss ≈ 0.03, Final validation loss ≈ 0.01(Fig. 20), theconfusion matrix for a classification model tested on the CASME II. The confusion matrix summarizes the model's performance across different classes.X-axis (Predicted labels): 0 to 25, Y-axis (True labels): 0 to 25, Cell values indicate the number of samples predicted as a class, darker cells indicate higher values (correct classifications are along the diagonal). Colourbar: Indicates frequency scale from 0 to 50+. (See fig. 21). The Receiver Operating Characteristic (ROC) curves for multiple classes in the CASME II dataset. X-axis has False Positive Rate (FPR) and Y-axis has True Positive Rate (TPR); all classes (except one) have an AUC of 1.00 indicating perfect classification performance. Each curve corresponds to a different class (SUB01 to SUB26), All valid ROC curves follow the top-left corner path (i.e., from (0, 0) to (0, 1) to (1, 1)). This is the ideal ROC shape, indicating zero false positives and perfect true positive detection. (Fig. 22).

Figure 18: Happiness-related expression from CASME II dataset

Table 3: Comparisons for the CASME II dataset

Methods	Accuracy (%)
FaceSleuth (SOA) [44]	95.1
FACS-Based Graph Features [46]	75.04
RCGA-CNN (optimized CNN) [52]	89.24
SlowFast CNN (Rank-1 identification) [53]	87.4
TSDN (Two-Stream Difference Network) [54]	~71.49

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Attribute-Embedding + Contrastive Learning (3D CNN + BERT) [55]	**77.82**
Vision-Transformer variant (convolutional patch embeddings + dual/attention branches) [61]	**99.00**
Proposed	**100.00**

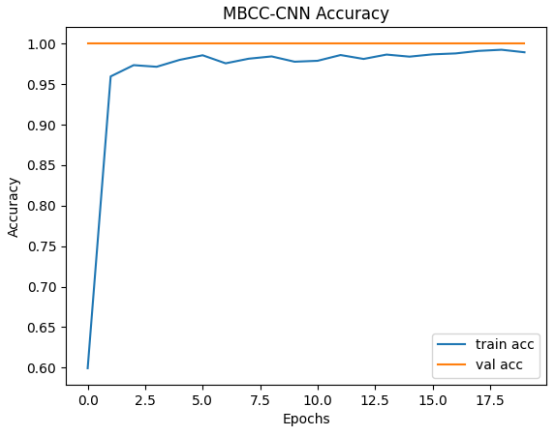

Figure 19: Training and validation accuracy curves for CASME II dataset

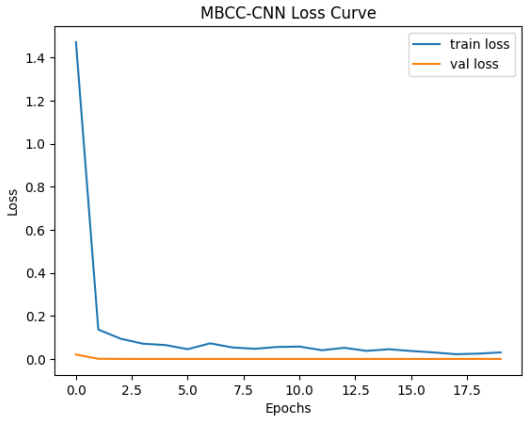

Figure 20: Loss curves for the MBCC-CNN model trained on the CASME II dataset

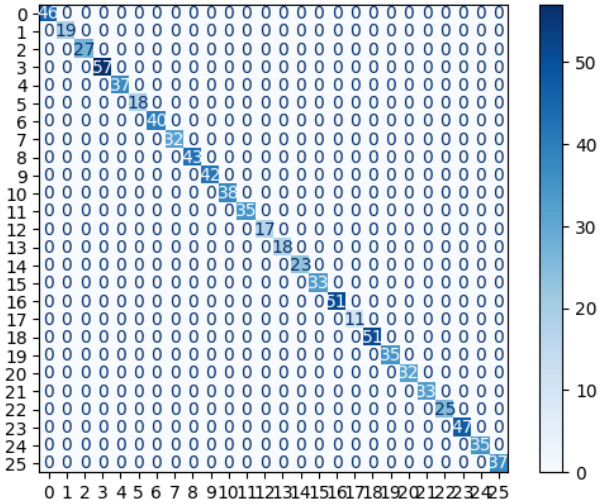

Figure 21: Confusion matrix for a classification model tested on the CASME II dataset

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

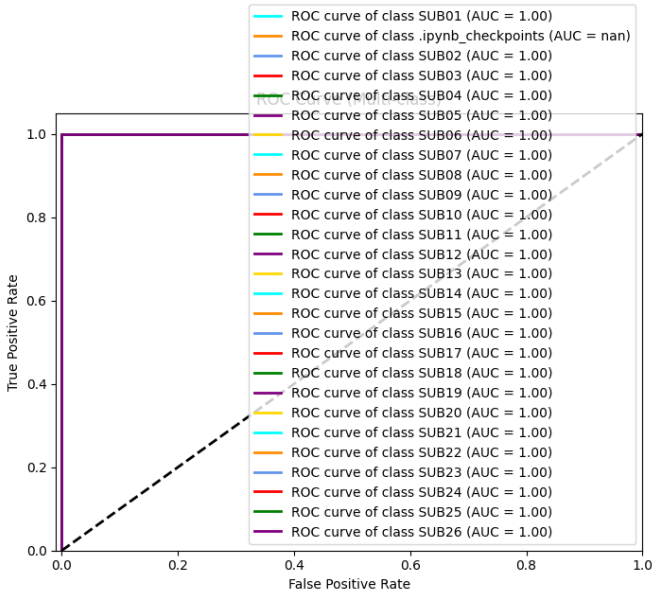

Figure 22: ROC curve for multiple classes in CASME II dataset

4.4 Spontaneous Macro-Expressions and Micro-Expressions (CAS(ME)2) Dataset Results

Qu et al. [72] developed a novel facial expression dataset incorporating both macro- and micro-expressions, comprising 250 and 53 samples respectively, selected from over 600 recorded facial movements. The dataset was built using recordings from 22 subjects (6 males and 16 females) with an average age of 22.59 years (SD = 2.2). The recordings were captured using a Logitech Pro C920 webcam at a resolution of 640 × 480 pixels with a frame rate of 30 fps. Annotation of the CAS(ME)² dataset was carried out by combining Action Units (AUs), participants' self-reports, and emotional categories determined from emotion-inducing videos. The dataset includes four classes of emotions: positive, negative, surprise, and others. Detailed analysis of the comparison table for the CAS(ME)2 dataset that includes performance (accuracy) of various state-of-the-art methods used for micro-expression recognition. The table shows significant differences in performance (see table 4). The training and validation accuracy curves of a model named MBCC-CNN trained on the CAS(ME)2 dataset, achieving a final accuracy of ~99.58%. X-axis has Epochs (0 to 19), Y-axis has Accuracy (from 0.5 to 1.0), Blue line shows Training accuracy (train accuracy), Orange line shows Validation accuracy (val accuracy). Training accuracy starts at ~0.48 and climbs sharply, Validation accuracy starts higher (~0.75), suggesting the model generalizes well early on. By epoch 5, both curves exceed 98% accuracy. From epoch 6 onward, both training and validation accuracies stabilize near 1.0, indicating Model convergence, High classification confidence, and Minimal loss over time (confirmed by your loss curve) (see fig. 23) and the loss curves for the MBCC-CNN model trained on the CAS(ME)2, showing how both training and validation loss evolves across epochs. The final loss reaches ~0.04, indicating very high model confidence and minimal error. This is a training vs validation loss curve for the MBCC-CNN model over 20 epochs, X-axis has Epochs (0–19), Y-axis has Loss (scale: 0 to ~1.1), Blue line depicts Training loss, orange line depicts Validation loss. Both training and validation loss decrease rapidly in the first 5 epochs. By around epoch 7, both losses plateau at a very low value (near zero).Training Loss Starts at ~1.05 and consistently drops, after epoch 7, it remains around 0.01 or lower, indicating excellent fit to training data and Validation Loss Starts around ~0.78 and drops in parallel with the training loss, Plateaus around ~0.02 after epoch 7.No signs of overfitting — validation loss doesn't increase or diverge after training loss bottoms out.(See fig. 24), the confusion matrix for a classification model tested on the CAS(ME)2. The confusion matrix summarizes the model's performance across different classes; this is a 3-class classification confusion matrix. The structure has Rows for Actual/True classes and Columns for Predicted classes. The cell at (i, j) indicates how many times samples of class i (actual) were predicted as class j (predicted). Class 0 (ANGER), Correctly classified are 353, Misclassified 1, Class 1 (DISGUST), Correctly classified are 232, No misclassifications, Class 2 (HAPPY), Correctly classified are 132, are samples, DISGUST (class 1) was classified with 100% accuracy — very distinctive features likely helped. (see fig. 25). The Receiver Operating Characteristic (ROC) curves for multiple classes in the

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CAS(ME)2 dataset. ROC curves are used to find the performance of classification models by showing the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR). AUC (Area under the Curve) Values are for ANGER it is AUC = 1.00, DISGUST it is AUC = 1.00, HAPPY it is AUC = 1.00, AUC = 1.00 means perfect classification for those classes. The model distinguishes perfectly between the given class and others (Fig. 26).

Figure 25: Sample images for CAS(ME)²

Table 4.Comparison table for the CAS(ME)² dataset

Methods	Accuracy (%)
3D-DenseNet + Squeeze-and-Excitation [62]	92.96
Hybrid Attention-3DNet (HA-3DNet) [63]	93.95
Optical-flow / spotting and recognition baselines (MEGC / MTSN related) [64]	~79.2
LBP-TOP baseline (various LBP-TOP parameter settings) [72]	75.66
3D SE-DenseNet (EVM + apex-based preprocessing + 3D SE-DenseNet) [63]	92.96
MADV-Net (multi-scale feature pyramid + dynamic attention / Transformer improvements) [73]	83.32
Proposed	**99.58**

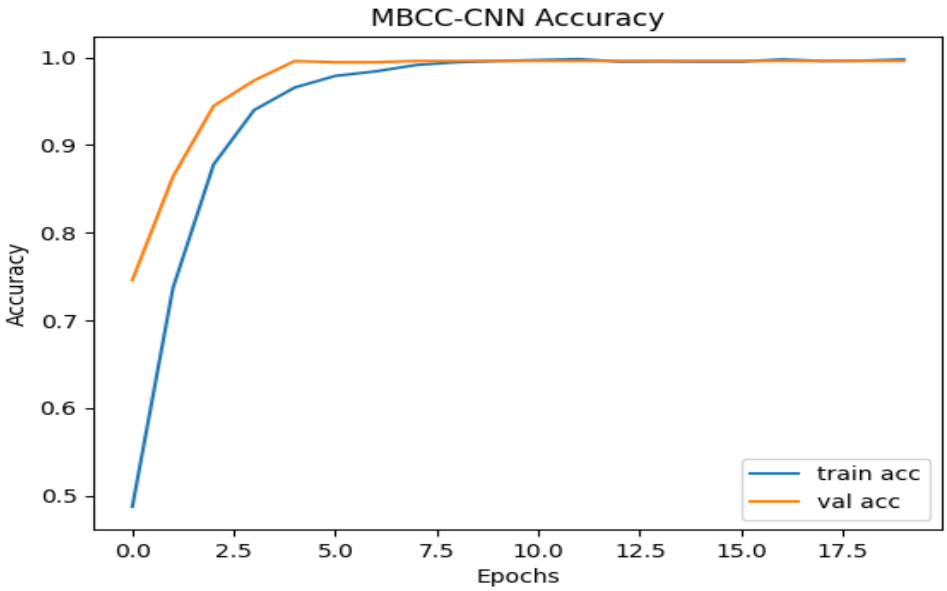

Figure 23: Training and validation accuracy curves for CAS(ME)²

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

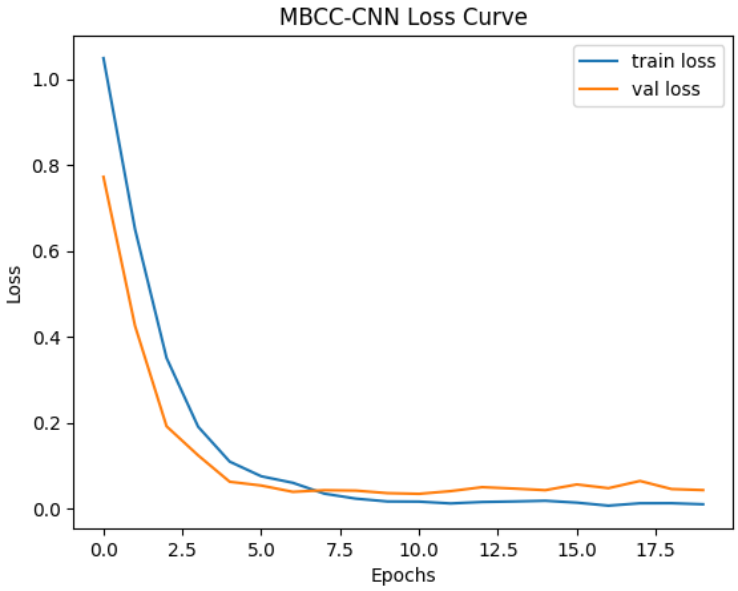

Figure 24: Loss curves for the MBCC-CNN model trained on the CAS(ME)² dataset

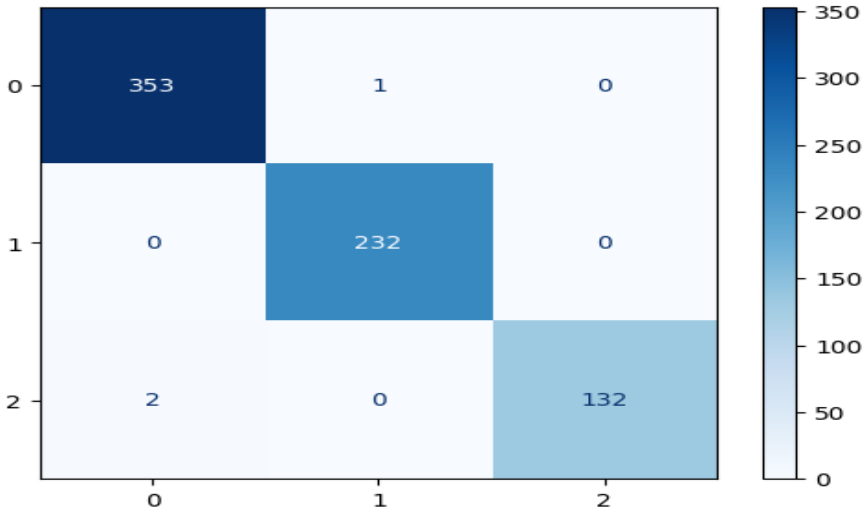

Figure 25: Confusion matrix for a classification model tested on the CAS(ME)²

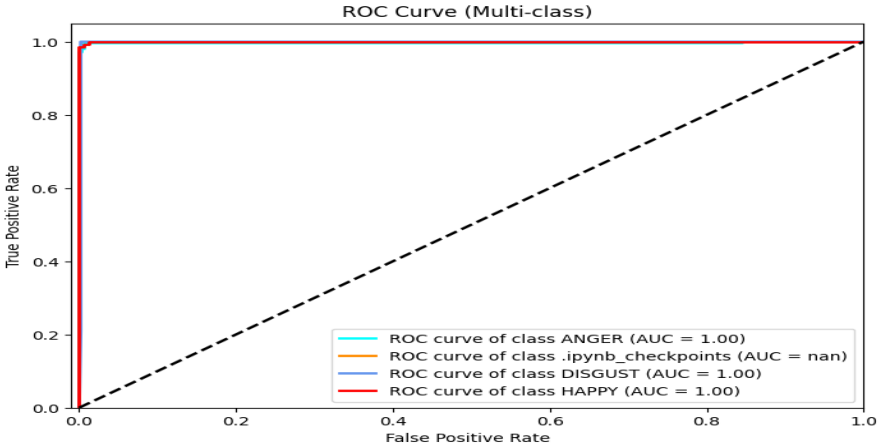

Figure 26: ROC curve for multiple classes in CAS(ME)² dataset

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CONCLUSIONS

In this work, we introduce a multi-branch cross-connection convolutional neural network (MBCC-CNN) framework for facial expression recognition, which integrates concepts from residual learning, Network-in-Network, and hierarchical multi-branch structures. The proposed approach begins with pre-processing the input facial images, followed by systematic extraction of discriminative features. These features are learned independently across multiple parallel branches, each focusing on different representational aspects, and later combined to enhance the overall feature learning capacity of the MBCC-CNN. To further refine the learned representations, global average pooling is employed on the final feature maps, and the outputs are directly fed into a SoftMax classifier for expression categorization.

Experimental evaluations demonstrate that the proposed MBCC-CNN consistently outperforms several existing techniques, showing improvements across multiple evaluation metrics and exhibiting greater robustness. Thereby facilitating intelligent and real-time deployment in practical scenarios. Addressing the challenge of achieving reliable facial expression recognition in complex, real-world environments remains an open research direction, which we intend to investigate in future studies.

REFERENCES

[1] K. R. Scherer, “What are emotions? And how can they be measured?” Soc. Sci. Inf., vol. 44, no. 4, pp. 695–729, 2005.ured?” Soc. Sci. Inf., vol. 44, no. 4, pp. 695–729, 2005
[2] A. Freitas-Magalh ~ aes, “The psychology of emotions: The allure of humanface,” Leya, 2020.
[3] C. A. Corneanu, M. O. Simon, J. F. Cohn, and S. E. Guerrero, “Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 8, pp. 1548–1568, Aug. 2016.
[4] P. Ekman, “Emotions revealed: Recognizing faces and feelings to improve communication and emotional life,” Holt Paperback, vol. 128, no. 8, pp. 140–140, 2003.
[5] P. Ekman and W. V. Friesen, “Nonverbal leakage and clues to deception,” Psychiatry-interpersonal Biol. Processes, vol. 32, no. 1, pp. 88–106, 1969.
[6] B. Bhushan, “Study of facial micro-expressions in psychology,” in Understanding Facial Expressions in Communication, Berlin, Ger many: Springer, 2015, pp. 265–286.
[7] W. J. Yan, Q. Wu, J. Liang, Y. H. Chen, and X. Fu, “How fast are the leaked facial expressions: The duration of micro expressions,” J. Nonverbal Behav., vol. 37, no. 4, pp. 217–230, 2013.
[8] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition). New York, NY, USA: WW Norton andCompany, 2009.
[9] S. Porter and L. T. Brinke, “Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions,” Psychol. Sci., vol. 19, no. 5, pp. 508–514, 2008.
[10] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan, “I see how you feel: Training laypeople and professionals to recognize f leeting emotions,” in Proc. Annu. Meeting Int. Commun. Assoc., 2013, pp. 3515–3522.
[11] S. Nag, A. K. Bhunia, A. Konwer, and P. P. Roy, “Facial micro expression spotting and recognition using time contrasted feature with visual memory,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. 2019, pp. 2022–2026.
[12] S. Li and W. Deng, Deep facial expression recognition: A sur vey, IEEE Trans. Affect. Comput., early access, Mar. 17, 2020, doi: 10.1109/TAFFC.2020.2981446.
[13] P. Liu, S. Han, Z. Meng, and Y. Tong, Facial expression recognition via a boosted deep belief network, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Columbus, OH, USA, Jun. 2014, pp. 18051812.
[14] Z. Yu and C. Zhang, Image based static facial expression recognition with multiple deep network learning, in Proc. ACMInt. Conf. Multimodal Interact. Washington, DC, USA, Nov. 2015, pp. 435442.
[15] A. Mollahosseini, D. Chan, and M. H. Mahoor, Going deeper in facial expression recognition using deep neural networks, in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2016, pp. 110.
[16] J.ShaoandY.Qian, Threeconvolutionalneuralnetworkmodelsforfacial expression recognition in the wild, Neurocomputing, vol. 355, pp. 8292, Aug. 2019.

[17] S. Xie, H. Hu, and Y. Wu, Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition, Pat tern Recognit., vol. 92, pp. 177191, Aug. 2019.
[18] D. K. Jain, P. Shamsolmoali, and P. Sehdev, Extended deep neural network for facial emotion recognition, Pattern Recognit. Lett. vol. 120, pp. 6974, Apr. 2019.
[19] P. Ekman, "An argument for basic emotions," Cognition and Emotion, vol. 6, pp. 169–200, 1992.
[20] Paul Ekman, Emotions Revealed: Understanding Faces and Feelings. Phoenix, 2004.
[21] P. Ekman and E. L. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), ser. Series in Affective Science. Oxford University Press, 2005.
[22] J. A. Russell and J. M. Fern´ andez-Dols, The psychology of facial expression. Cambridge university press, 1997.
[23] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," Psychiatry, vol. 32, no. 1, pp. 88–106, 1969.
[24] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3d-gradient descrip tor," in Crime Detection and Prevention (ICDP 2009), 3rd International Conference on. IET, 2009, pp. 1–6.
[25] M. G. Frank, C. J. Maccario, and V. l. Govindaraju, "Behavior and security," in Protecting airline passengers in the age of terrorism. Greenwood Pub. Group, 2009.
[26] Y.-l. Tian, T. Kanade, and J. F. Cohn, Evaluation of Gabor-wavelet based facial action unit recognition in image sequences of increasing complexity, in Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit., Washington, DC, USA, May 2002, pp. 229234.
[27] L. Zhong, Q. Liu, P. Yang, J. Huang, and D. N. Metaxas, Learning multiscale active facial patches for expression analysis, IEEE Trans. Cybern., vol. 45, no. 8, pp. 14991510, Aug. 2015.
[28] R. Girshick, J. Donahue, T. Drrell, and J. Malik, Rich feature hierarchies for accurate object detection and demantic segmentation, in Proc. IEEE Conf. Comput. Vis. Pattern Recogit., Jun. 2014, pp. 580587.
[29] Truong Le Vinh, Phucand Tien, Le and Tri, Duong. (2024). Facial Expression Recognition using Traditional Machine Learning Models.
[30] Liu, X. (2024). The application of machine learning and deep learning based algorithms in facial expression recognition. In Proceedings of the International Conference on Emerging Trends in Machine Intelligence and Technology (EMITI). SciTePress.
[31] The first look: A biometric analysis of emotion recognition using key facial features. (2025). Frontiers in Computer Science.
[32] Kumar, T. A., et al. (2024, October 2). Enhancing facial emotion level recognition: A CNN based approach to balancing data. In Proceedings of the International Conference on Advanced Information and Communication Technology (AICTC). Springer.
[33] Alkhalelī, N. F. A., andBecerikli, Y. (2024, July 31). Facial expression recognition and emotion detection with CNN methods and SVM classifiers. Journal of Millimeterwave Communication, Optimization and Modelling.
[34] El Boudouri, Y., andBohi, A. (2025, January 14). EmoNeXt: An adapted ConvNeXt for facial emotion recognition. arXiv preprint arXiv:2501.06543.
[35] X. Sun, P. Xia, L. Zhang, and L. Shao, A ROI-guided deep architecture for robust facial expressions recognition, Inf. Sci., vol. 522, pp. 3548, Jun. 2020.
[36] S. Minaee and A. Abdolrashidi, Deep-emotion: Facial expression recognition using attentional convolutional network, 2019, arXiv:1902.01019. [Online]. Available: http://arxiv.org/abs/1902.01019
[37] X. Sun, S. Zheng, and H. Fu, ROI-attention vectorized CNN model for static facial expression recognition, IEEE Access, vol. 8, pp. 71837194, 2020.
[38] Y.Gan,J.Chen,Z.Yang,andL.Xu, Multipleattention network for facial expression recognition, IEEE Access, vol. 8, pp. 73837393, 2020.
[39] Rashad, M., et al. (2024, January). CCNN SVM: Automated model for emotion recognition based on custom convolutional neural networks with SVM. Electronics. MDPI.
[40] From constricted models to state of the art for facial expression classification. (2025, July 15). SN Computer Science. Springer.

[41] Yu, J., Zheng, Y., Wang, L., Wang, Y., andXu, S. (2025, March 15).Design of an expression recognition solution employing the global channel spatial attention mechanism. arXiv preprint arXiv:2503.09876.
[42] A. Veit, M. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in Proc. Adv. Neural Inf. Process. Syst., May 2016, pp. 1–9.
[43] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, arXiv:1312.4400. [Online]. Available: https://arxiv.org/abs/1312.4400
[44] Wu, L., Jiang, T., Duan, W., Fang, Y., and Keung, J. (2025). FaceSleuth: Learning-Driven Single-Orientation Attention Verifies Vertical Dominance in Micro-Expression Recognition. arXiv preprint arXiv:2506.02695.
[45] Shi, C., Tan, C., and Wang, L. (2021). A facial expression recognition method based on a multibranch cross-connection convolutional neural network. IEEE access, 9, 39255-39274.
[46] Buhari, A. M., Ooi, C. P., Baskaran, V. M., Phan, R. C., Wong, K., and Tan, W. H. (2020). FACS-based graph features for real-time micro-expression recognition. Journal of Imaging, 6(12), 130.
[47] Shi, H., Wang, Y., Wang, R., and Liu, D. (2025). AUMEs: AU Detection-Based Dual-Stream Multi-task 3DCNN for Micro-expression Recognition. Neural Processing Letters, 57(1), 18.
[48] Sie-Min, K., Zulkifley, M. A., and Kamari, N. A. M. (2022). Optimal compact network for micro-expression analysis system. Sensors, 22(11), 4011.
[49] Gan, Y. S., Liong, S. T., Yau, W. C., Huang, Y. C., and Tan, L. K. (2019). OFF-ApexNet on micro-expression recognition system. Signal Processing: Image Communication, 74, 129-139.
[50] 梁詩婷. A Shallow Triple Stream Three-dimensional CNN (STSTNet) for Micro-expression Recognition System.
[51] Wei, M., Zong, Y., Jiang, X., Lu, C., and Liu, J. (2022). Micro-expression recognition using uncertainty-aware magnification-robust networks. Entropy, 24(9), 1271.
[52] Naidana, K. S., Yarra, Y., andDivvela, L. P. (2025). Facial micro-expression classification through an optimized convolutional neural network using genetic algorithm. Bulletin of Electrical Engineering and Informatics, 14(1), 307-315.
[53] Kay, T., Ringel, Y., Cohen, K., Azulay, M. A., andMendlovic, D. (2023). Person Recognition using Facial Micro-Expressions with Deep Learning. arXiv preprint arXiv:2306.13907.
[54] Pan, H., Xie, L., Li, J., Lv, Z., and Wang, Z. (2021). Micro-expression recognition by two-stream difference network. IET Computer Vision, 15(6), 440-448.
[55] Song, Y., Wang, J., Wu, T., Huang, Z., and Xiao, J. (2022, July). Micro-expression recognition based on attribute information embedding and cross-modal contrastive learning. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.
[56] Verma, M., Vipparthi, S. K., Singh, G., andMurala, S. (2019). LEARNet: Dynamic imaging network for micro expression recognition. IEEE Transactions on Image Processing, 29, 1618-1627.
[57] Wang, Y., Huang, Y., Liu, C., Gu, X., Yang, D., Wang, S., and Zhang, B. (2021). Micro Expression Recognition via Dual-Stream Spatiotemporal Attention Network. Journal of Healthcare Engineering, 2021(1), 7799100.
[58] Esmaeili, V., MohasselFeghhi, M., andShahdi, S. O. (2022). Automatic micro-expression recognition using LBP-SIPl and FR-CNN. AUT Journal of Modeling and Simulation, 54(1), 59-72.
[59] Cen, S., Yu, Y., Yan, G., Yu, M., and Yang, Q. (2020). Sparse spatiotemporal descriptor for micro-expression recognition using enhanced local cube binary pattern. Sensors, 20(16), 4437.
[60] Reddy, G. V., Reddy, S. P. T., Mukherjee, S., andDubey, S. R. (2021, January). MERANet: facial micro-expression recognition using 3D residual attention network. In ICVGIP.
[61] Indolia, S., Nigam, S., Singh, R., Singh, V. K., and Singh, M. K. (2023). Micro expression recognition using convolution patch in vision transformer. IEEE Access, 11, 100495-100507.
[62] Cai, L., Li, H., Dong, W., and Fang, H. (2022). Micro-expression recognition using 3D DenseNet fused Squeeze-and-Excitation Networks. Applied Soft Computing, 119, 108594.
[63] Irawan, B., Munir, R., Utama, N. P., andPurwarianti, A. (2025). ENHANCING MICRO-EXPRESSION RECOGNITION: A NOVEL APPROACH WITH HYBRID ATTENTION-3DNET. Jordanian Journal of Computers and Information Technology, 11(1).

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[64] Jingting, L. I., Wang, S. J., Yap, M. H., See, J., Hong, X., and Li, X. (2020, November). Megc2020-the third facial micro-expression grand challenge. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (pp. 777-780). IEEE.

[65] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "Samm: A spontaneous micro-facial movement dataset," IEEE Transactions on Affective Computing, vol. 9, no. 1, pp. 116–129, Jan 2018.

[66] Emotion, vol. 6, pp. 169–200, 1992. Paul Ekman, Emotions Revealed: Understanding Faces and Feelings. Phoenix, 2004.

[67] expression. Cambridge university press, 1997. P. Ekman, "Lie catching and microexpressions," in The Philosophy of Deception, C. W. Martin, Ed. Oxford University Press, 2009, pp. 118–133.

[68] Merghani, W., Davison, A. K., and Yap, M. H. (2018). A review on facial micro-expressions analysis: datasets, features and metrics. arXiv preprint arXiv:1805.02397.

[69] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, "Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces," in Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013, pp. 1–7.

[70] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "Casme ii: An improved spontaneous micro-expression database and the baseline evaluation," PloS one, vol. 9, no. 1, 2014.

[71] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press, 1978.

[72] Qu, F., Wang, S. J., Yan, W. J., and Fu, X. (2016, June). CAS (ME) 2: A Database of Spontaneous Macro-expressions and Micro-expressions. In International Conference on Human-Computer Interaction (pp. 48-59). Cham: Springer International Publishing.

[73] Kong, W., You, Z., andLv, X. (2025). 3D Micro-Expression Recognition Based on Adaptive Dynamic Vision. Sensors, 25(10), 3175.